REPORT DOCUMENTATION PAGE

AFRL-SR-BL-TR-00-

0493

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED 01 Aug 97 - 01 Nov 99

4. TITLE AND SUBTITLE "Machine Learning via Mathematical Programming	5. FUNDING NUMBERS F49620-97-1-0326
6. AUTHOR(S) Olivi L. Mamgasarian	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) The Board of Regents of University of Wisconsin 750 University Avenue Madison, WI 53706-1490	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) AFOSR 801 North Randolph Street, Room 732 Arlington, VA 22203-1977	10. SPONSORING/MONITORING AGENCY REPORT NUMBER F49620-97-1-0326

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION AVAILABILITY STATEMENT Approved for public release, distribution unlimited	12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Mathematical programming approaches were applied to a variety of problems in machine learning in order to gain deeper understanding of the problems and to come up with new and more efficient computational algorithms. Theoretical and/or computational contributions were made to Data Envelopment Analysis wherein one seeks efficient decision making units, Neural Networks with as few hidden units as possible, optimization problems subject to constraints that in turn require the solution of further optimization problems, classification algorithms that suppress unnecessary or redundant features, algorithms that "chunk" massive data sets in order to classify them, clustering data based on the novel concept of nearness to cluster planes rather than cluster centroids, a new implementable general theory for Support Vector Machines that does away with the restrictive Mercer positive definite kernel condition that had hitherto been universally assumed, a very effective Successive Over relaxation (SOR) algorithm for solving very large linear and nonlinear kernel classification problems, applying support vector machines to breast cancer diagnosis and prognosis, smoothing algorithms for solving large and complex classification problems, nonlinear data fitting using support vector machines and a robust loss function, and classifying data that is partly labeled and partly unlabeled.

20001002 016

14. SUBJECT TERMS			15. NUMBER OF PAGES 8
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT

Approved 20 Sep 00 RCH

From: Olvi Mangasarian [olvi@cs.wisc.edu]
Sent: Tuesday, September 19, 2000 1:48 PM
To: rob.herklotz@afosr.af.mil
Cc: olvi@cs.wisc.edu
Subject: Revised Final Report for Grant F49620-97-1-0326

1. Cover Sheet

Final Report for AFOSR Grant F49620-97-1-0326 (%%%%%Revised Sep 19, 2000%%%%%)
"Machine Learning via Mathematical Programming"
PI: Olvi L. Mangasarian
Institution: University of Wisconsin - Madison
Reporting Period: 08-01-97 - 11-01-99

2. Objectives

Apply mathematical programming to solve machine learning and related problems such as data mining and knowledge discovery.

2a. Abstract of Results

Mathematical programming approaches were applied to a variety of problems in machine learning in order to gain deeper understanding of the problems and to come up with new and more efficient computational algorithms. Theoretical and/or computational contributions were made to Data Envelopment Analysis wherein one seeks efficient decision making units, Neural Networks with as few hidden units as possible, optimization problems subject to constraints that in turn require the solution of further optimization problems, classification algorithms that suppress unnecessary or redundant features, algorithms that "chunk" massive datasets in order to classify them, clustering data based on the novel concept of nearness to cluster planes rather than cluster centroids, a new implementable general theory for Support Vector Machines that does away with the restrictive Mercer positive definite kernel condition that had hitherto been universally assumed, a very effective Successive Overrelaxation (SOR) algorithm for solving very large linear and nonlinear kernel classification problems, applying support vector machines to breast cancer diagnosis and prognosis, smoothing algorithms for solving large and complex classification problems, nonlinear data fitting using support vector machines and a robust loss function, and classifying data that is partly labeled and partly unlabeled.

3. Research

The research supported by this grant resulted in:

(a) Seventeen papers, most of which are already published in refereed journals or conference proceedings. These papers are listed in Section 5 and are easily available on the web as indicated by the links given in Section 5 following each paper.

(b) Twenty talks given at 15 national and international meetings, workshops and at universities.

3a. Summary of Results (Numbers refers to Section 5 below)

In (5(i)) we consider the problem of projecting a point in a polyhedral set onto the boundary of the set using an arbitrary norm for the projection.
Two types of polyhedral sets, one defined by a convex combination of k points in R^n and the second by the intersection of m closed halfspaces in R^n, lead to disparate optimization problems for finding such a projection. The first case leads to a mathematical program with a linear

objective function and constraints that are linear inequalities except for a single nonconvex cylindrical constraint. The second polyhedral set leads to a much simpler problem of determining the minimum of m easily evaluated numbers. Similarly disparate mathematical programs ensue from the problem of finding the largest ball relative to the affine hull of a polyhedral set, with radius measured by an arbitrary norm, that can be inscribed in the polyhedral set. For a polyhedral set of the first type this problem leads to a maxmin of a bilinear function over linear inequality constraints and a single nonconvex cylindrical constraint, while for the second type this problem leads to a single linear program. Interestingly, for the one norm, the nonconvex mathematical program associated with the boundary projection problem for the first polyhedral set can be solved by solving 2n linear programs.

In (5(ii)) a fast parsimonious linear-programming-based algorithm for training neural networks is proposed that suppresses redundant features while using a minimal number of hidden units. This is achieved by propagating sideways to newly added hidden units the task of separating successive groups of unclassified points. Computational results show an improvement of 26.53% and 19.76% in tenfold cross-validation test correctness over a parsimonious perceptron on two publicly available datasets.

In (5(iii)) we consider an arbitrary linear program with equilibrium constraints (LPEC) that may possibly be infeasible or have an unbounded objective function. We regularize the LPEC by perturbing it in a minimal way so that the regularized problem is solvable. We show that such regularization leads to a problem that is guaranteed to have a solution which is an exact solution to the original LPEC if that problem is solvable, otherwise it is a residual-minimizing approximate solution to the original LPEC. We propose a finite successive linearization algorithm for the regularized problem that terminates at point satisfying the minimum principle necessary optimality condition for the problem.

In (5(iv)) an overview of the rapidly emerging research and applications area of data mining is given. In addition to providing a general overview, motivating the importance of data mining problems within the area of knowledge discovery in databases, our aim is to list some of the pressing research challenges, and outline opportunities for contributions by the optimization research communities. Towards these goals, we include formulations of the basic categories of data mining methods as optimization problems. We also provide examples of successful mathematical programming approaches to some data mining problems.

In (5(v)) computational comparison is made between two feature selection approaches for finding a separating plane that discriminates between two point sets in an n-dimensional feature space selecting as few of the n features (dimensions) as possible. In the concave minimization approach a separating plane is generated by minimizing a weighted sum of distances of misclassified points to two parallel planes that bound the sets and which determine the separating plane midway between them. Furthermore, the number of dimensions of the space used to determine the plane is minimized. In the support vector machine approach, in addition to minimizing the weighted sum of distances of misclassified points to the bounding planes, we also maximize the distance between the two bounding planes that generate the separating plane. Computational results show that feature suppression is an indirect consequence of the support vector machine approach when an appropriate

norm is used.
Numerical tests on 6 public data sets show that classifiers trained by the concave minimization approach and those trained by a support vector machine have comparable 10-fold cross-validation correctness. However, in all data sets tested, the classifiers obtained by the concave minimization approach selected fewer problem features than those trained by a support vector machine.

In (5(vi)) a linear support vector machine formulation is used to generate a fast, finitely-terminating linear-programming algorithm for discriminating between two massive sets in n-dimensional space, where the number of points can be orders of magnitude larger than n. The algorithm creates a succession of sufficiently small linear programs that separate chunks of the data at a time. The key idea is that a small number of support vectors, corresponding to linear programming constraints with positive dual variables, are carried over between the successive small linear programs, each of which containing a chunk of the data. We prove that this procedure is monotonic and terminates in a finite number of steps at an exact solution that leads to a globally optimal separating plane for the entire dataset. Numerical results on fully dense publicly available datasets, numbering 20,000 to 1 million points in 32-dimensional space, confirm the theoretical results and demonstrate the ability to handle very large problems.

In (5(vii)) a finite new algorithm is proposed for clustering m given points in n-dimensional real space into k clusters by generating k planes that constitute a local solution to the nonconvex problem of minimizing the sum of squares of the 2-norm distances between each point and a nearest plane. The key to the algorithm lies in a formulation that generates a plane in n-dimensional space that minimizes the sum of the squares of the 2-norm distances to each of m_1 given points in the space. The plane is generated by an eigenvector corresponding to a smallest eigenvalue of an n-by-n simple matrix derived from the m_1 points. The algorithm was tested on the publicly available Wisconsin Breast Prognosis Cancer database to generate well separated patient survival curves. In contrast, the k-mean algorithm did not generate such well-separated survival curves.

In (5(viii)) by setting apart the two functions of a support vector machine: separation of points by a nonlinear surface in the original space of patterns, and maximizing the distance between separating planes in a higher dimensional space, we are able to define indefinite, possibly discontinuous, kernels, not necessarily inner product ones, that generate highly nonlinear separating surfaces. Maximizing the distance between the separating planes in the higher dimensional space is surrogated by support vector suppression, which is achieved by minimizing any desired norm of support vector multipliers. The norm may be one induced by the separation kernel if it happens to be positive definite, or a Euclidean or a polyhedral norm. The latter norm leads to a linear program whereas the former norms lead to convex quadratic programs, all with an arbitrary separation kernel. A standard support vector machine can be recovered by using the same kernel for separation and support vector suppression. On a simple test example, all models perform equally well when a positive definite kernel is used. When a negative definite kernel is used, we are unable to solve the nonconvex quadratic program associated with a conventional support vector machine, while all other proposed models remain convex and easily generate a surface that separates all given points.

In (5(ix)) successive overrelaxation (SOR) for symmetric linear complementarity problems and quadratic programs is used to train a support vector machine (SVM) for discriminating between the elements of two massive datasets, each with millions of points. Because SOR handles one point at a time, similar to Platt's sequential minimal optimization (SMO) algorithm which handles two constraints at a time, it can process very large datasets that need not reside in memory. The algorithm converges linearly to a solution. Encouraging numerical results on very large datasets that cannot be processed by conventional linear or quadratic programming methods are presented.

In (5(x)) we show that new formulations of support vector machines can generate nonlinear separating surfaces which can discriminate between elements of a given set better than a linear surface. The principal approach used is that of generalized support vector machines (GSVMs) which employ possibly indefinite kernels. The GSVM training procedure is carried out by either the simple successive overrelaxation (SOR) iterative method or by linear programming. This novel combination of powerful support vector machines with the highly effective SOR computational algorithm or with linear programming allows us to use a nonlinear surface to discriminate between elements of a dataset that belong to one of two categories. Numerical results on a number of datasets show improved testing set correctness, by as much as a factor of two, when comparing the nonlinear GSVM surface to a linear separating surface.

In (5(xi)) we define prognostic relationships between computer-derived nuclear morphological features, lymph node status, and tumor size in breast cancer. Computer-derived nuclear size, shape and texture features were determined from fine-needle aspirates obtained at the time of diagnosis from 253 consecutive patients with invasive breast cancer. Tumor size and lymph node status were determined at the time of surgery. If our relationships by others, axillary dissection for breast cancer staging, estimating prognosis, and selecting patients for adjunctive therapy could be eliminated.

In (5(xii)) we describe the role of generalized support vector machines in separating massive and complex data using arbitrary nonlinear kernels. Feature selection that improves generalization is implemented via an effective procedure that utilizes a polyhedral norm or a concave function minimization. Massive data is separated using a linear programming chunking algorithm as well as a successive overrelaxation algorithm, each of which is capable of processing data with millions of points.

In (5(xiii)) the problem of tolerant data fitting by a nonlinear surface, induced by a kernel-based support vector machine , is formulated as a linear program with fewer number of variables than that of other linear programming formulations. A generalization of the linear programming chunking algorithm for arbitrary kernels is implemented for solving problems with very large datasets wherein chunking is performed on BOTH data points and problem variables. The proposed approach tolerates a small error, which is adjusted parametrically, while fitting the given data. This leads to improved fitting of noisy data as demonstrated computationally. Comparative numerical results indicate an average time reduction as high as 26.0%, with a maximal time reduction of 79.7%. Additionally, linear programs with as many as 16,000 data points and more than a billion nonzero matrix elements are solved.

In (5(xiv)) smoothing methods, extensively used for solving important mathematical programming problems and applications, are applied here to generate and solve an unconstrained smooth reformulation of the support vector machine for pattern classification using a completely arbitrary kernel. We term such reformulation a smooth support vector machine (SSVM). A fast Newton-Armijo algorithm for solving the SSVM converges globally and quadratically. Numerical results and comparisons are given to demonstrate the effectiveness and speed of the algorithm. On six publicly available datasets, tenfold cross validation correctness of SSVM was the highest compared with four other methods as well as the fastest. On larger problems, SSVM was comparable or faster than SVM^light SOR and SMO. SSVM can also generate a highly nonlinear separating surface such as a checkerboard.

In (5(xv)) we show that Kernel Principal Component Analysis (KPCA) has proven to be a versatile tool for unsupervised learning, however at a high computational cost due to the dense expansions in terms of kernel functions. We overcome this problem by proposing a new class of feature extractors employing l_1 norms in coefficient space instead of the reproducing kernel Hilbert space in which KPCA was originally formulated in. Moreover, the modified setting allows us to efficiently extract features maximizing criteria other than the variance much in a projection pursuit fashion.

In (5(xvi)) a mixed integer programming semisupervised support vector machine (S^3VM) proposed by Bennett and Demiriz for classification of partially labeled two-class datasets, is trained here as a concave S^3VM (VS^3VM) using a very fast finitely terminating successive linear programming algorithm that can handle much larger unlabeled datasets than the mixed integer programming approach.
For partially labeled datasets the algorithm assigns unlabeled data to one of two classes so as to maximize the separation between the two classes.
For labeled data the testing set part of the data is treated as unlabeled data in VS^3VM.
For unlabeled data, a k-median clustering algorithm is used to select a small percentage, say 10%, to be labeled by an expert or an oracle. This labeled set is used together with the remaining part of the data, that remains unlabeled, in VS^3VM. Numerical testing indicate a relative test set improvement, as high as 20%, over a standard supervised linear programming procedure that is trained on a randomly chosen set that is labeled and used as a training set.

In (5(xvii)) the robust Huber M-estimator, a differentiable cost function that is quadratic for small errors and linear otherwise, is modeled exactly by an easily solvable simple convex quadratic program for both linear and nonlinear support vector estimators. In contrast, all previous models involved specialized numerical algorithms for solving the robust Huber linear estimator. Numerical test comparisons with these algorithms indicate the computational effectiveness of the new quadratic programming model for both linear and nonlinear support vector problems. Results are shown on problems with as many as 20,000 data points, with considerably faster running times on larger problems.

4. Personnel Supported

Paul S. Bradley, Ph.D. Candidate. Degree granted August 1998.
David Musicant, Ph.D. Candidate. Degree expected August 2000.
Yuh-Jye Lee, Ph.D. Candidate. Degree expected August 2001.

5. Publications

(i) O. L. Mangasarian: ``Polyhedral boundary projection", Mathematical Programming Technical Report 97-10, October 1997, SIAM Journal on Optimization, 9, 1999, 1128-1134. ftp://ftp.cs.wisc.edu/math-prog/tech-reports/97-10.ps

(ii) P. S. Bradley and O. L. Mangasarian: ``Parsimonious side propagation",

ICASSP98: IEEE International Conference on Acoustics, Speech and Signal Processing, Seattle May 12-15, 1998, Volume 3, 1873-1876. ftp://ftp.cs.wisc.edu/math-prog/tech-reports/97-11.ps

(iii) O. L. Mangasarian: ``Regularized linear programs with equilibrium constraints", in ``Reformulation-Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods", M. Fukushima and Liqun Qi, editors, Kluwer Academic Publishers, 1998, 259-268. ftp://ftp.cs.wisc.edu/math-prog/tech-reports/97-13.ps

(iv) P. S. Bradley, Usama M. Fayyad and O. L. Mangasarian: ``Data mining: overview and optimization opportunities", Mathematical Programming Technical Report 98-01, January 1998, INFORMS Journal on Computing 11, 1999, 217-238. ftp://ftp.cs.wisc.edu/math-prog/tech-reports/98-01.ps

(v) P. S. Bradley and O. L. Mangasarian:``Feature selection via concave minimization and support vector machines", in ``Machine Learning Proceedings of the Fifteenth International Conference (ICML '98)", Madison, WI, July 24-27, 1998, Morgan Kaufmann, San Francisco, CA 1998, 82-90 ftp://ftp.cs.wisc.edu/math-prog/tech-reports/98-03.ps

(vi) P. S. Bradley and O. L. Mangasarian: ``Massive data discrimination via linear support vector machines", Mathematical Programming Technical Report 98-05, May 1998. Optimization Methods and Software, 13(1), 2000, 1-10. ftp://ftp.cs.wisc.edu/math-prog/tech-reports/98-05.ps

(vii) P. S. Bradley and O. L. Mangasarian: ``k-Plane Clustering", Mathematical Programming Technical Report 98-08, August 1998. Journal of Global Optimization 16, Number 1, 2000, 23-32. ftp://ftp.cs.wisc.edu/math-prog/tech-reports/98-08.ps

(viii) O. L. Mangasarian: ``Generalized Support Vector Machines", Mathematical Programming Technical Report 98-14, October 1998. ``Advances in Large Margin Classifiers", A. J. Smola, P. Bartlett, B. Sch\"{o}kopf and D. Schuurmans, editors, MIT Press, 1999, 135-146. ftp://ftp.cs.wisc.edu/math-prog/tech-reports/98-14.ps

(ix) O. L. Mangasarian and D. R. Musicant: ``Successive Overrelaxation for Support Vector Machines", Mathematical Programming Technical Report 98-18, November 1998, IEEE Transactions on Neural Networks 10, 1999, 1032-1037. ftp://ftp.cs.wisc.edu/math-prog/tech-reports/98-18.ps

(x) O. L. Mangasarian and D. R. Musicant: ``Data Discrimination via Nonlinear Generalized Support Vector Machines", Mathematical Programming Technical Report 99-03, March 1999. ``Applications and Algorithms of Complementarity", M. C. Ferris, O. L. Mangasarian and J.-S. Pang, editors, Kluwer Academic Publishers, 2000, to appear. ftp://ftp.cs.wisc.edu/math-prog/tech-reports/99-03.ps

(xi) W. H. Wolberg, W. N. Street and O. L. Mangasarian: ``Importance of Nuclear Morphology in Breast Cancer Prognosis", Clinical Cancer Research 5, 1999, 3542-3548.

(xii) P. S. Bradley, O. L. Mangasarian and D. R. Musicant: ``Optimization Methods in Massive Datasets", Data Mining Institute Technical Report 99-01, June 1999. ``Handbook of Massive Datasets", J. Abello , P. M. Pardalos, M. G. C. Resende, editors, Kluwer Academic Publishers 2000, to appear. ftp://ftp.cs.wisc.edu/pub/dmi/tech-reports/99-01.ps

(xiii) O. L. Mangasarian and D. R. Musicant: ``Massive Support Vector Regression", Data Mining Institute Technical Report 99-02, August, 1999. NIPS*99 Workshop on Learning with Support Vectors: Theory and Applications. Machine Learning, submitted. ftp://ftp.cs.wisc.edu/pub/dmi/tech-reports/99-02.ps

(xiv) Y.-J. Lee and O. L. Mangasarian: ``SSVM: A Smooth Support Vector Machine for Classification", Data Mining Institute Technical Report 99-03, September 1999. Computational Optimization and Applications, to appear. ftp://ftp.cs.wisc.edu/pub/dmi/tech-reports/99-03.ps

(xv) A. J. Smola, O. L. Mangasarian and B. Sch\"olkopf: ``Sparse Kernel

Feature Analysis", Data Mining Institute Technical Report 99-04, October 1999. Neural Computation, submitted.
ftp://ftp.cs.wisc.edu/pub/dmi/tech-reports/99-04.ps

(xvi) G. Fung and O. L. Mangasarian: ``Semi-Supervised Support Vector Machines for Unlabeled Data Classification", Data Mining Institute Technical Report 99-05, October 1999. Optimization Methods and Software, submitted.
ftp://ftp.cs.wisc.edu/pub/dmi/tech-reports/99-05.ps

(xvii) O. L. Mangasarian and D. R. Musicant: ``Robust Linear and Support Vector Regression", Data Mining Institute Technical Report 99-09, November 1999.
IEEE Transactions on Pattern Analysis and Machine Intelligence, accepted.
ftp://ftp.cs.wisc.edu/pub/dmi/tech-reports/99-09.ps

6. Interactions/Transitions
----

a. Meetings, Conferences & Seminars

(I) International Symposium on Mathematical Programming, Lausanne, Switzerland, August 24-29, 1997

(i) Talk: "Feature selection by mathematical programming"
(ii) Talk: "Minimum-support solutions of mathematical programs"
(iii) Talk: "Data mining via concave minimization"

(II) West Coast Optimization Meeting, Departments of Mathematics and Applied Mathematics, University of Washington, Seattle, WA, November 14-15, 1997.

Talk: "Data mining via bilinear programming"

(III) INFORMS National Meeting, Montreal, Quebec, April 26-29, 1998

Talk: "Minimum-support solutions for the ill-posed linear complementarity problem"

(IV) IEEE International Conference on Acoustics, Speech and Signal Processing, Seattle, WA, May 12-15, 1998

Talk: "Parsimonious side propagation"

(V) Seminar, University of California at San Diego, June 16, 1998

Talk: "Massive data discrimination via linear support vector machines"

(VI) International Conference on Machine Learning, Madison, WI, July 23-26, 1998--
----

Talk: "Feature Selection via Concave Minimization and Support Vector Machines"

(VII) INFORMS National Meeting, Seattle, WA, October 25-28, 1998

(i) Talk: "Polyhedral Boundary Projection"
(ii) Talk: "Breast Cancer Prognosis without Lymph Node Status"

(VIII) NIPS*98 Workshop on Large Margin Classifiers, Breckenridge, CO, December 4-5, 1998.

(i) Talk: "Mathematical Programming in Machine Learning"
(ii)Talk: "Successive Overrelaxation for Support Vector Machines"

(IX) Seminar, University of California, San Diego, January 12, 1999

Talk: "Successive Overrelaxation for Support Vector Machines"

(X) AFOSR Meeting, Air Force Academy, Colorado Springs, CO, February 3-4, 1999

Talk:"Massive Data Discrimination via Support Vector Machines"

(XI) Invited Plenary Talk to Joint Annual SIAM Meeting and the Optimization Conference, Atlanta, May 10-12, 1999.

Talk: "Optimization in Machine Learning and Data Mining"

(XII) International Conference on Complementarity, Madison, June 9-12, 1999

Talk:"Nonlinear Data Discrimination via Generalized Support Vector Machines"

(XIII) INFORMS Fall 99 Conference, Philadelphia, November 7-9, 1999

(i) Talk:"Smoothing Methods for Support Vector Machines"
(ii) Talk:"Generalized Support Vector Machines for Data Discrimination"

(XIV) NIPS 1999 "Learning with Support Vector Machines: Theory and Applications", Workshop, Breckenridge, CO, December 2-4, 1999

Talk:"Massive Support Vector Regression"

(XV) DIMACS Workshop on Discrete Mathematical Problems and Medical Applications, DIMACS Center, Rutgers University, Piscataway, NJ, December 8-10, 1999

Talk: "Breast Cancer Survival Analysis and Chemotherapy via Generalized Support Vector Machines"

b. Transitions

XCYT, our linear-programming-based non-invasive breast cancer diagnostic system, continues to be used and improved upon at University Hospital, with a very high accuracy.

c. Mentions in the Media

(i) Marilynn Marchione: ``Detecting Changes in Breast Cancer Diagnosis", Milwaukee Sentinel, October 10, 1999.
www.cs.wisc.edu/~olvi/media/MilwSent.html

(ii) ``Operations Research: The Science and Technology for Informed Decision Making" , National ITV Satellite Schedule, Wednesday October 13, 1999, 1430 - 1500 ET and Wednesday November 17, 1430 - 1500 ETr, Channel 513.

(iii) James Case: ``Data Mining Emerges as a New Discipline in a World of Increasingly Massive Data Sets", SIAM News, Volume 32, Number 10, pages 1 \& 4, December 1999.
www.cs.wisc.edu/~olvi/media/mining.pdf